

Research Article

Comparison of Machine Learning Methods for the Arterial Hypertension Diagnostics

Vladimir S. Kublanov,¹ Anton Yu. Dolganov,¹ David Belo,² and Hugo Gamboa²

¹Research Medical and Biological Engineering Centre of High Technologies, Ural Federal University, Mira 19, Yekaterinburg 620002, Russia

²Laboratório de Instrumentação, Engenharia Biomédica e Física da Radiação (LIBPhys-UNL), Departamento de Física, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Monte da Caparica, 2892-516 Caparica, Portugal

Correspondence should be addressed to Vladimir S. Kublanov; kublanov@mail.ru

Received 4 April 2017; Revised 29 May 2017; Accepted 1 June 2017; Published 31 July 2017

Academic Editor: Justin Keogh

Copyright © 2017 Vladimir S. Kublanov et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The paper presents results of machine learning approach accuracy applied analysis of cardiac activity. The study evaluates the diagnostics possibilities of the arterial hypertension by means of the short-term heart rate variability signals. Two groups were studied: 30 relatively healthy volunteers and 40 patients suffering from the arterial hypertension of II-III degree. The following machine learning approaches were studied: linear and quadratic discriminant analysis, k -nearest neighbors, support vector machine with radial basis, decision trees, and naive Bayes classifier. Moreover, in the study, different methods of feature extraction are analyzed: statistical, spectral, wavelet, and multifractal. All in all, 53 features were investigated. Investigation results show that discriminant analysis achieves the highest classification accuracy. The suggested approach of noncorrelated feature set search achieved higher results than data set based on the principal components.

1. Introduction

According to the World Health data, hypertension affects more than 1 billion people worldwide. Many factors can conduce to hypertension, including occipital stress and job strain [1]. One of the main problems concerning the treatments for the arterial hypertension conditions is the late detection for apparently healthy people. Some studies had been shown that among the individuals with hypertension, more than 35% were unaware of their condition [2].

The heart rate variability (HRV) is among one of the widely used biomedical signals, due to ease of record the electrical heart activity [3]. The HRV analysis can be applied in the task of arterial hypertension diagnostics, since it is well known that various features of the HRV reflect behavior of the different modules of the autonomic nervous system (ANS) [4].

Common HRV analysis implies the application of a variety of analysis methods: statistical, spectral, and nonlinear

analysis. Generally, in a single study, a limited number of features are extracted. Such as in [5], 13 nonlinear features were studied for efficacy of stress state detection. The current paper comprises a study of 53 different features. Usually during one study, feature sets of a particular method are used. For example in [6], sets of time-domain features, nonlinear features, and spectral features were studied separately for automatic sleep staging by means of HRV signal analyses. In this study, combinations of different methods were analyzed.

The common uses of machine learning approaches for condition classification based on HRV information imply usage of several available methods: support vector machine (SVM), discriminant analysis (DA), and ordinal pattern statistics (OPS) [7–9]. However, the selection of a particular approach is not always justified. In current study, linear and quadratic DA, SVM, k -nearest neighbors, decision trees, and Naïve Bayes approaches were investigated.

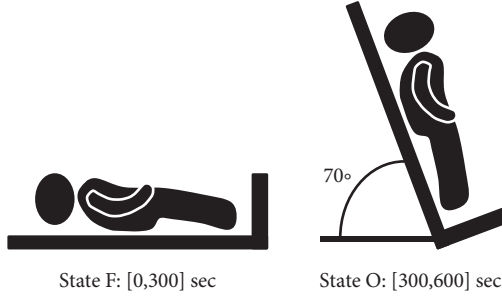


FIGURE 1: Diagram of the study.

In one of the previous works, the investigation of the linear and quadratic discriminant analysis was carried out, implying the study of arterial hypertension diagnostic using single features of short-term HRV signals. In that work, the evaluation of the features and the evaluation of the classifier efficacy were carried out by means of an in-house software produced in MATLAB [10]. In the present paper, the machine learning methods were implemented in the python.

In summary, the goal of the present work is to study the efficacy of different machine learning approaches for diagnostic of the arterial hypertension by means of the short-term HRV, using combinations of statistical, spectral (Fourier and Wavelet transforms), and nonlinear features. By applying feature combinations of different methods, we aim to build more robust and accurate classifiers.

2. Materials and Methods

2.1. Recorded Dataset. The clinical part of the study was performed in the Sverdlovsk Clinical Hospital of Mental Diseases for Military Veterans (Yekaterinburg, Russian Federation). For the HR signals registration, the electroencephalograph-analyzer “Encephalan-131-03” (“Medicom-MTD,” Taganrog, Russian Federation) was used. The rotating table Lojer (Vammalan Konepaja OY, Finland) performed the spatial position change of the patient during passive orthostatic load; the lift of the head end of the table was up to 70° from the horizontal position. The clinical part of the study was approved by the local Ethics Committee of the Ural State Medical University.

Participants of this study were 30 healthy volunteers and 41 patients suffering from the arterial hypertension of II and III degree. The electrocardiography (ECG) signals were recorded in two functional states: functional rest (state F) and passive orthostatic load (state O). The length of the signal in the mentioned state was about 300 seconds. The HRV signals were consequently derived from the ECG signals automatically by the “Encephalan-131-03” software. Figure 1 presents diagram of the functional states.

2.2. Heart Rate Variability Features. Prior to the processing, the original time series were cleaned from the artifacts. By artifacts, in this study, we considered values of the R-R

intervals that differed from the HR mean by more than three standard deviations. NN is the abbreviation for the “normal to normal” time series, that is, without artifacts. Among all studied time series, less than 2% of data was removed. For spectral and multifractal analyses, NN time series were interpolated using cubic spline interpolation with the 10 Hz sampling frequency.

The feature dataset is the same as used previously [11], where it was shown that features of the HRV signals recorded in the state O have better classification accuracy for arterial hypertension diagnostics. Therefore, in this study, we analyze data only in the state O. The used features were separated into statistical, geometric, spectral (Fourier based), wavelet, nonlinear, and multifractal. Their description will be given below.

2.2.1. Statistical Features. Statistical methods are used for the direct quantitative evaluation of the HR time series. Main quantitative features are as follows:

- (i) M is the mean value of the R-R intervals after artifact rejection:

$$M = \frac{1}{N} \sum_{i=1}^N NN_i, \quad (1)$$

where N is the number of elements in the NN and NN_i is the i th element in R-R time series.

- (ii) HR, the heart rate, is an inverse ratio to M :

$$HR = \frac{1}{M}. \quad (2)$$

- (iii) SDNN is the standard deviation of the NN intervals:

$$SDNN = \sqrt{\frac{1}{N} - 1 \sum_{i=1}^N (NN_i - M)^2}. \quad (3)$$

- (iv) CV, the coefficient of variation, is defined as ratio of standard deviation SDNN to the mean M , expressed in percent;

$$CV = \frac{SDNN}{M} \cdot 100\%. \quad (4)$$

- (v) RMSSD is the square root of mean of squares of differences between successive elements in NN;

$$RMSSD = \left[\frac{1}{N} \sum_{i=1}^{N-1} (NN_{i+1} - NN_i)^2 \right]^{0.5}. \quad (5)$$

- (vi) NN50 is the number of pairs of successive elements in NN that differ by more than 50 ms [12].

2.2.2. Geometric Features. The geometric methods analyze the distribution of the R-R intervals as a random numbers. The common features of these methods are as follows:

- (i) M_0 , the mode, is the most frequent value in the R-R interval. In case of the normal distribution, the mode is close to the mean M .
- (ii) VR, the variation range, is the difference between the lowest R-R interval and the highest R-R interval in the time series. VR shows variability of the R-R interval values and reflects activity of the parasympathetic system of the ANS.
- (iii) AM_0 , the amplitude of the mode, is a number of the R-R intervals that correspond to the mode value. AM_0 shows the stabilizing effect of the heart rate management, mainly caused by the sympathetic activity [12].

The following indexes are derived from common geometric features:

- (i) SI, the stress index, reflects centralization degree of the heart rate and mostly characterizes the activity of the sympathetic department of the ANS:

$$SI = \frac{AM_0}{2M_0 \cdot VR}. \quad (6)$$

- (ii) IAB, the index of the autonomic balance, depends on the relation between activities of the sympathetic and parasympathetic department of the ANS:

$$IAB = \frac{AM_0}{VR}. \quad (7)$$

- (iii) ARI, the autonomic rhythm index, shows parasympathetic shifts of the autonomic balance: smaller values of the ARI correspond to the shift of the autonomic balance to the parasympathetic activity:

$$ARI = \frac{1}{M_0 \cdot VR}. \quad (8)$$

- (iv) IARP, the index of adequate regulation processes, reflects accordance of the autonomic function changes of the sinus node as a reaction of the sympathetic regulatory effects on the heart:

$$IARP = \frac{AM_0}{M_0}. \quad (9)$$

2.2.3. Spectral Features. Spectral analysis is used to quantify periodic processes in the heart rate by the means of the Fourier transform (Fr). The main spectral components of the HRV signal are high frequency—HF (0.4–0.15 Hz), low frequency—LF (0.15–0.04 Hz), very low frequency—VLF

(0.04–0.003 Hz), and ultralow frequency—ULF (lower than 0.003 Hz) [12, 13]. For smaller than 300 seconds, short-term time series ULF spectral component is not analyzed.

HF spectral component characterizes activity of the parasympathetic system of the ANS and activity of the autonomic regulation loop. High frequencies of the heart rate in HRV spectrum are associated with the breathing and determined by the connection and influences of the vagus with the sinus node.

LF spectral component mainly characterizes activity of the sympathetic vascular tone regulation center. Low frequencies reflect modulation of the heart rate by the sympathetic nervous system [4].

VLF spectral component is defined by the suprasegmental regulation of the heart rate, as the amplitude of the VLF waves and is related to the psycho-emotional strain and functional state of the cortex. The genesis of the very low frequencies is still the matter of debates. Most likely, it is influenced by the suprasegmental centers of the autonomic regulation that generates slow rhythms. These rhythms are directed to the heart by the sympathetic nervous system, humoral factors on the sinus node. Biologic rhythms in the same frequency band are connected with the mechanisms of the thermoregulation, fluctuations of the vascular tone, the renin activity, and the secretion of the leptin [14]. The similarity of the frequencies implies the participation of these mechanisms in the genesis of the VLF spectral component. There are evidences of the increase of the VLF activity in case of the central nervous system damage, anxiety, and depression disorders [15].

The studied quantitative features of spectral analysis are

- (i) spectral power of the HF, LF, and VLF components,
- (ii) total power of the spectrum—TP,
- (iii) normalized values of the spectral components by the total power— HF_n , LF_n , and VLF_n ,

$$\begin{aligned} HF_n &= \frac{HF}{TP}, \\ LF_n &= \frac{LF}{TP}, \\ VLF_n &= \frac{VLF}{TP}, \end{aligned} \quad (10)$$

- (iv) the LF/HF ratio, also known as the autonomic balance exponent,
- (v) IC, the index of centralization,

$$IC = \frac{HF + LF}{VLF}, \quad (11)$$

- (vi) IAS, the index of the subcortical nervous centers activation,

$$IAS = \frac{LF}{VLF}, \quad (12)$$

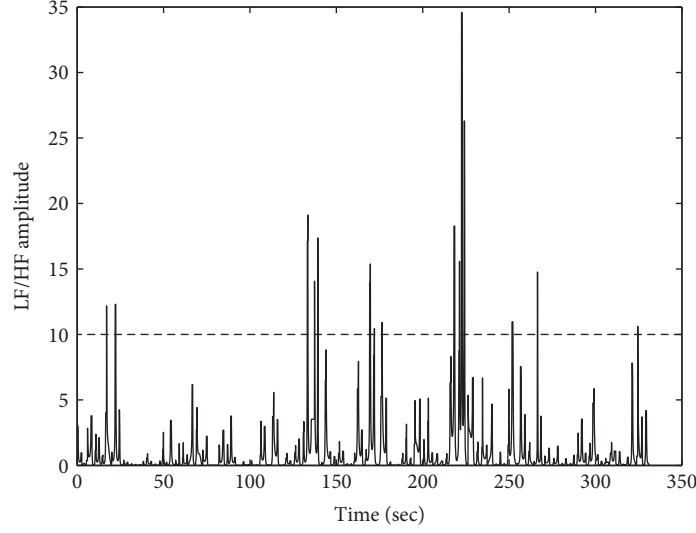


FIGURE 2: Example of $F[LF_{wt}/HF_{wt}(t)]$ with the decision threshold.

- (vii) HF_{max} , the maximal power of the HF spectral components,
- (viii) RF, the respiration frequency, frequency that corresponds to the HF_{max} [16].

2.2.4. Wavelet Transform. For nonstationary time series, one can also use the wavelet transform (wt), to simultaneously study time-frequency patterns. The general equation for continuous wavelet transform is as follows:

$$W(a, b) = \frac{1}{\sqrt{a}} \int s(t) \cdot \psi\left(\frac{t-b}{a}\right) dt, \quad (13)$$

where a is the scale, b is the shift, ψ is the wavelet basis, and $s(t)$ is the analyzed signal [17].

Moreover, the connection between the scale and the analyzed frequency is in accordance with the following:

$$a = \frac{f_c * f_s}{f}, \quad (14)$$

where f_c is the central frequency of the wavelet basis, called by the *centfrq* function, f_s is the sampling frequency of the analyzed signal, and f is the analyzed frequency. For wavelet transform computation in this work, we used wavelet Coiflet of the fifth order [18].

It is possible to acquire same spectral features by means of the wavelet transform:

- (i) Spectral power of the HF, LF, and VLF components
- (ii) Normalized values of the spectral components by the total power— HF_n , LF_n , and VLF_n
- (iii) The LF/HF ratio.

Additionally, standard deviations $SDHF(wt)$, $SDLF(wt)$, and $SDVLF(wt)$ of the $HF_{wt}(t)$, $LF_{wt}(t)$, and $VLF_{wt}(t)$ time

series were tested as features. $HF_{wt}(t)$, $LF_{wt}(t)$, and $VLF_{wt}(t)$ are time series of the HF, LF, and VLF spectral components, respectively, acquired by means of the wavelet transform.

Moreover, one can study informational characteristics of the wavelet transform by analyzing the $F[LF_{wt}/HF_{wt}(t)]$ function (Figure 2). $F[LF_{wt}/HF_{wt}(t)]$ is the continuous function of the LF/HF ratio. This function was not a smooth morphology. Its “excursions” (local dysfunctions) varies in case of functional loads, as the features of $F[LF_{wt}/HF_{wt}(t)]$ is possible to use the number of dysfunctions N_d , the maximal value of dysfunction $(LF/HF)_{max}$, and the intensity of dysfunction $(LF/HF)_{int}$. By the dysfunction, we consider values of function that suppress decision threshold Δ according to previous studies of our research group $\Delta = 10$ [19].

2.2.5. Nonlinear Feature. As the nonlinear feature in this study, we have used the Hurst exponent calculated by the aggregated variance method. The variance can be written as follows:

$$\text{Var}|(X(t_2) - X(t_1))^2| = \sigma^2 |t_2 - t_1|^{2H}, \quad (15)$$

where H is the Hurst exponent and X is a time-series vector.

H can be defined as the slope exponent in the following equation:

$$\log \sigma_{rms}(\Delta X) = c + H \log |s|, \quad (16)$$

where $\sigma_{rms}(\Delta X)$ is the standard deviation of the ΔX increments, corresponding to the time period s , and c is a constant [20].

Note that $H > 0.5$ corresponds to the process with trend, so-called persistent process, contrary $H < 0.5$ corresponds to antipersistent processes that have a tendency for trend change, and $H = 0.5$ is the random process [21].

2.2.6. Multifractal Features. As nonlinear methods, we adopted the multifractal detrended fluctuation analysis (MFDFA) [22]. The algorithm and application features of

the MFDFA method to estimation of short-term time series are described in details in [23].

The main steps of the method include the following:

- (i) The detrending procedure with second degree polynomial on nonoverlapping segments where the length of the segments corresponds to the studied time scale boundaries.

In current study, we investigated time scale boundaries that correspond to the LF and VLF frequency bands: 6–25 sec and 25–300 sec, respectively. In our earlier works and by other authors, it was noted that multifractal analysis of the HF component is not informative because of the noising [24].

- (ii) Determination of the fluctuation functions for q in range $q = [-5, 5]$:

$$F_x(q, s) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} \left[\frac{1}{s} \sum_{k=1}^s [\text{NN}(k) - \text{NN}_v(k)]^2 \right]^{q/2} \right\}^{1/q}, \quad (17)$$

where NN_v is the local trend in the segment v , N_s is the number of segments, and s is the scale.

- (iii) Estimation of the slope exponent H_x in log-log plot of the fluctuation function against scale s for each q :

$$F_{x(q,s)} \approx s^{H_x(q)}. \quad (18)$$

- (iv) Calculation of the scaling exponent $\tau(q)$:

$$\tau(q) = q \cdot H_x(q) - 1. \quad (19)$$

- (v) The Legendre transform application for the probability distribution of the spectrum estimation:

$$D(\alpha) = q \cdot \alpha - \tau. \quad (20)$$

Figure 3 represents the main features of the multifractal spectrum estimated by the MFDFA method. Here, H_0 is the height of the spectrum and represents the most probable fluctuations in the investigated time scale boundary of the signal; H_2 is the generalized Hurst exponent (also known as correlation degree); α_{\min} represents behavior of the smallest fluctuations in the spectrum; α_{\max} represents behavior of the greatest fluctuations in the spectrum; and $W = \alpha_{\max} - \alpha_{\min}$ is the width of multifractal spectrum that shows the variability of fluctuations in the spectrum. Multifractal characteristics are quantitative measures of the self-similarity and may characterize functional changes in the regulatory processes of the organism. In addition, we also tested the so-called 1/2-width measure of the spectrum, which is defined as

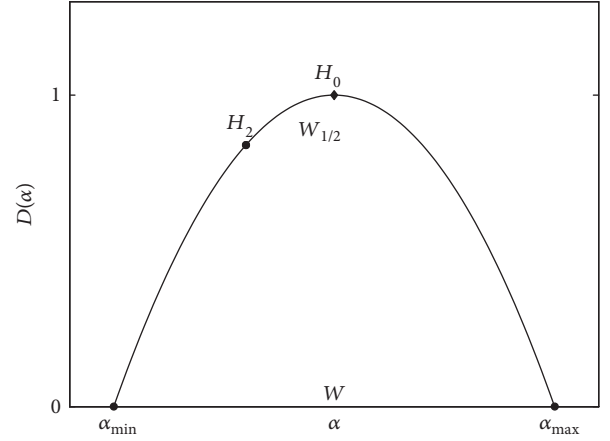


FIGURE 3: The features of multifractal analysis.

$W_{1/2} = |H_2 - H_0|$ [25]. Table 1 presents summary of all features used in this study.

2.3. Machine Learning Approaches. For the machine learning evaluation, the respective functions of the *sklearn* library were used [26]. The current paper describes supervised machine learning methods used. At first, the classifiers are trained using a training dataset. For that dataset, the class labels are known. After that, the efficacy of the classification is evaluated using test set of data. The efficacy is evaluated by comparing true labels of test set with those predicted by the model.

2.3.1. Discriminant Analysis (DA). In this work, two variants of the discriminant analysis were tested—linear and quadratic discriminant analyses (LDA and QDA). The LDA aims to find the best linear combination of the input features to properly separate studied classes. In the case of the QDA, the studied classes are separated by a quadratic function [27].

2.3.2. k -Nearest Neighbors (kNN). The k -nearest neighbors is one of the nonparametric machine learning approaches. In order to predict the class of the object, method chose the class, which is the most common among k “neighbors” of the object. Examples of the “neighbors” are picked from the training dataset. In the present study, different values of the k are tested—3, 4, and 5 [28].

2.3.3. Support Vector Machine (SVM). The base idea of the support vector machine methods is creation of the decision hyperplane which would separate different classes. In that case, the margin between two nearest points on the different sides of the hyperplane is maximal. In present study, the radial basis function (RBF) is used. For implementation in *python*, one have to specify the following: SVC(gamma = 2, C = 1) [29].

2.3.4. Decision Trees (DT). The decision trees classification model is built around a sequence of the Boolean queries. The sequence of such queries forms the “trees” structure. In the present work, variations of the classifier were analyzed—with fixed value of the maximal tree depth

TABLE 1: List of studied features.

Feature	Description	Equation
M	Mean value of the R-R	(1)
HR	Heart rate	(2)
SDNN	Standard deviation of the R-R	(3)
CV	Coefficient of the variation	(4)
RMSSD	Square root of mean of squares of differences between successive R-R	(5)
NN50	Variation higher than 50 ms in R-R signal	—
M_0	Mode of the R-R signal	—
VR	Variation range of the R-R signal	—
AM_0	Amplitude of the mode	—
SI	Stress index	(6)
IAB	Index of autonomic balance	(7)
ARI	Autonomic rhythm index	(8)
IARP	Index of adequate regulation processes	(9)
HF(Fr)	High frequency Fourier spectral power	—
LF(Fr)	Low frequency Fourier spectral power	—
VLF(Fr)	Very low frequency Fourier spectral power	—
TP(Fr)	Total power of the Fourier spectrum	—
LF/HF(Fr)	Autonomic balance exponent of the Fourier spectrum	—
$HF_{\max}(\text{Fr})$	Maximum power of the HF	—
$HF_n(\text{Fr}), LF_n(\text{Fr}), \text{ and } VLF_n(\text{Fr})$	Normalized power of the HF, LF, and VLF Fourier spectrum	(10)
IC	Index of centralization	(11)
IAS	Index of the subcortical nervous center's activation	(12)
RF	Respiration frequency	—
HF(wt)	High frequency wavelet spectral power	—
LF(wt)	Low frequency wavelet spectral power	—
VLF(wt)	Very low frequency wavelet spectral power	—
$HF_n(\text{wt}), LF_n(\text{wt}), \text{ and } VLF_n(\text{wt})$	Normalized power of the HF, LF, and VLF wavelet spectrum	—
SDHF(wt), SDLF(wt), and SDVLF(wt)	Standard deviations of the HF(t), LF(t), and VLF(t) wavelet time series	—
TP(wt)	Total power of the wavelet spectrum	—
LF/HF(wt)	Autonomic balance exponent of the wavelet spectrum	—
$(LF/HF)_{\max}$	Maximal value of dysfunctions	—
$(LF/HF)_{\text{int}}$	Intensity of dysfunctions	—
Nd	Number of dysfunctions	—
H	Hurst exponent	(13)
$\alpha_{\min} \text{ LF}, \alpha_{\min} \text{ VLF}$	Smallest fluctuations of the LF and VLF spectral band	—
$\alpha_{\max} \text{ LF}, \alpha_{\max} \text{ VLF}$	Greatest fluctuations of the LF and VLF spectral band	—
WLF, WVLF	Spectrum width of the LF and VLF spectral band	—
$H_2\text{LF}, H_2\text{VLF}$	Correlation degree of the LF and VLF spectral band	—
$H_0\text{LF}, H_0\text{VLF}$	Spectrum height of the LF and VLF spectral band	—
$W_{1/2}\text{LF}, W_{1/2}\text{VLF}$	1/2-width measure of the LF and VLF spectral band	—

($\max_depth=5$). The maximal depth feature points the maximal number of queries that is allowed to use before reaching leaf. The leaf node is the node that has no “children” [30].

2.3.5. Naive Bayes (NB). This method is based on the application of the Bayes' theorem with assumptions that data has

strong (or naive) independence. In current study, the Gaussian distribution of data is assumed [31].

2.4. Semi-optimal Search of the Noncorrelated Feature Space

2.4.1. Feature Set Selection. In the current investigation, all possible combinations of all features were analyzed. However, it is well known that using combined correlated features

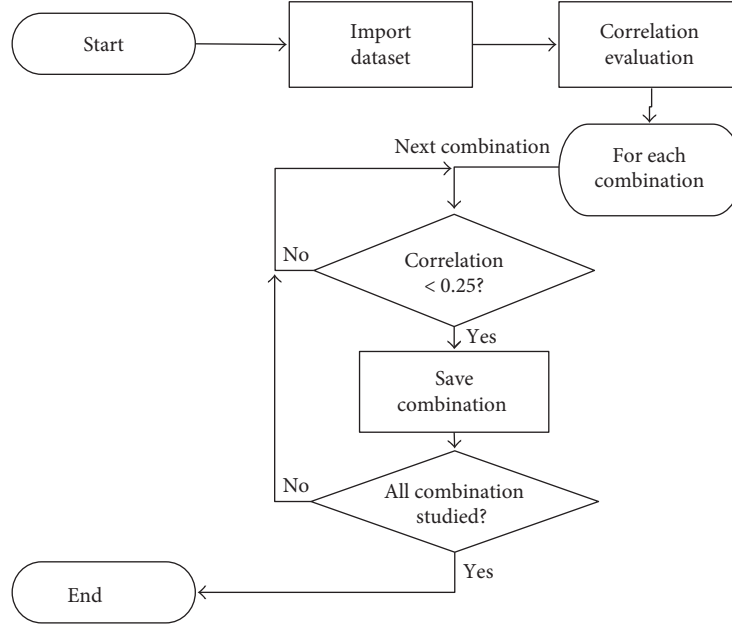


FIGURE 4: Flowchart of the noncorrelated combination selection.

in machine learning may lead to misleading results. Therefore, the first step in this investigation is to sort uncorrelated combinations. For this task, we compute the correlation coefficient. The whole flowchart of the script for noncorrelated feature combination selection is presented in Figure 4.

The threshold correlation value was set to 0.25. Usually, correlation more than 0.75 is considered to be high. Therefore, a value lower than 0.25 is a good benchmark for low correlation. In the current work, two to five feature combinations were made. In case of more than two feature combinations, the correlation was checked pairwise. When all calculation was finished, the noncorrelated features were saved to a file for future purposes.

Table 2 presents the total number of n -combinations for 53 features in case of $n = [2, 3, 4, 5]$ and number of selected noncorrelated combinations. Table 2 shows that application of such selection leads to both, more appropriate results and significant reduction of the analyzed combinations set.

2.4.2. Cross-Validation. Figure 5 presents a complete flowchart of the implemented algorithm for classifier efficacy evaluation.

Cross-validation implies division of the original datasets into m subsets, when $m-1$ subsets are used for the classifier training. The remaining part is used for the classifier test. The procedure is repeated m times. Such approach allows one to use dataset evenly [32].

In the current investigation, the number of random folds 1 was set to be 5. For the implementation of 5-fold cross-validation, we randomly divide the original dataset into 5 subsets. The division is implemented for both groups simultaneously. As the result, each subset included 6 healthy volunteers and 8 patients diagnosed with hypertension.

TABLE 2: Noncorrelated combination selection data.

n	Total n -combination number	Selected n -combinations	Calculation time, sec
2	1378	586	0.027
3	23,426	1669	0.477
4	292,825	1339	11.559
5	2,869,685	295	228.267

Many machine learning methods are sensitive to train set selection, so, in order to remove such influence, the cross-validation procedure was repeated 100 times with different folds. The repeated cross-validation allows to increase number of classification accuracy estimates [33].

Table 3 presents calculation times spent for each machine learning approach for different number of features in combinations. Calculation times are presented for all noncorrelated combinations. In accordance with Table 3, the fastest approach is the decision trees. The k -nearest neighbors approach is the slowest one.

3. Results

The classifier performance was averaged over 5 cross-validations and over 100 implementations. Figures 6–9 show overview of the classifier performance for all combinations for different numbers of features in combination. All color bars on Figures 5–8 have the same range—from 50 to 100%.

Figure 10 presents maximal accuracy achieved by each classifier for different number of features in set.

According to the data presented in Figures 5–9, the highest classification is achieved by the discriminant

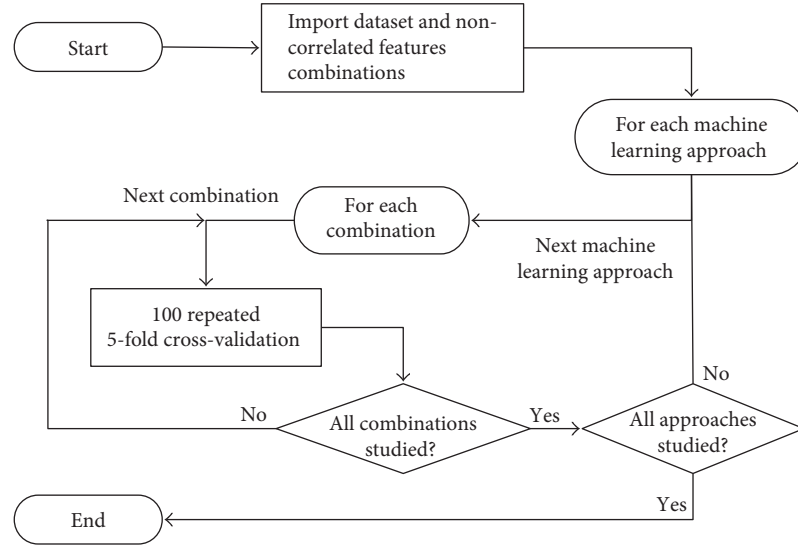


FIGURE 5: Flowchart of classifier efficacy evaluation algorithm.

TABLE 3: Calculation times of classifier efficacy evaluation, sec.

Features in combinations	LDA	QDA	NN3	NN4	NN5	RBF SVM	DT	Naive Bayes
2	165	113	281	281	281	139	89	130
3	482	346	917	860	806	403	269	375
4	397	288	640	643	642	325	222	300
5	88	64	141	140	140	71	50	66

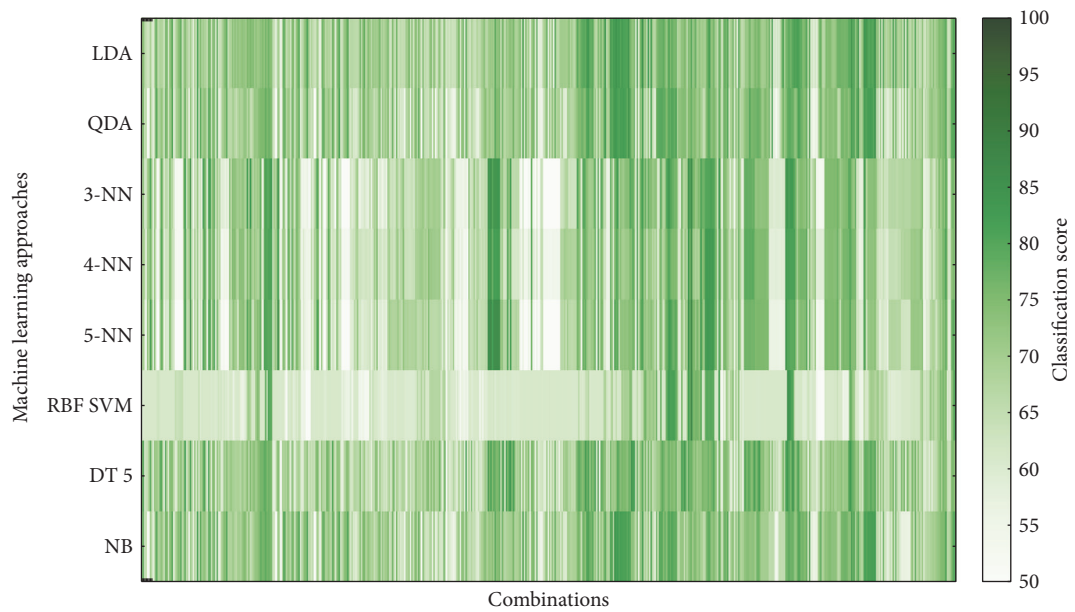


FIGURE 6: Classifier score for 2-feature combinations.

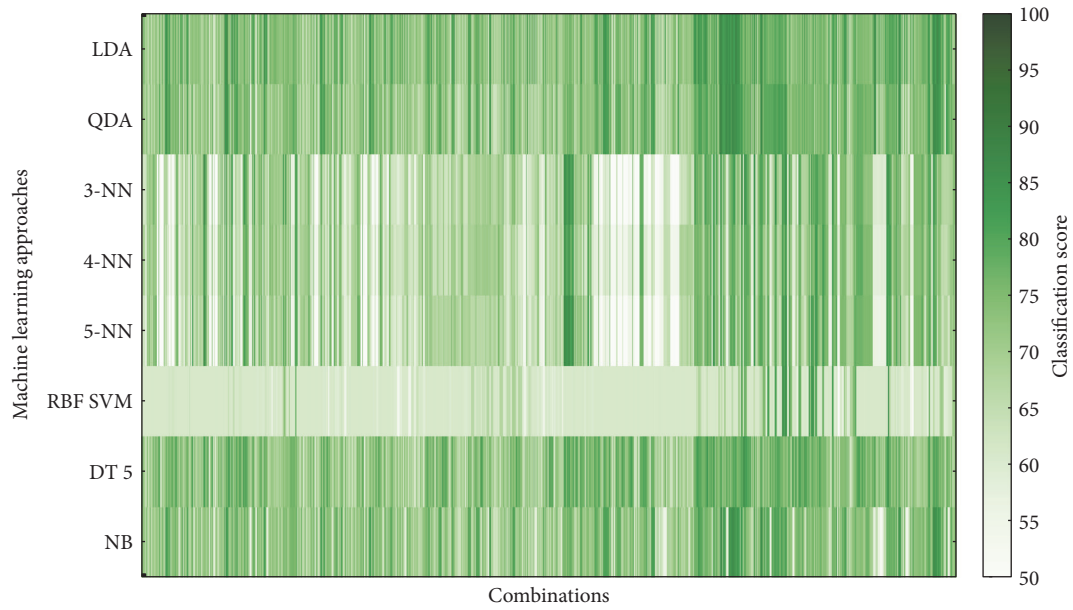


FIGURE 7: Classifier score for 3-feature combinations.

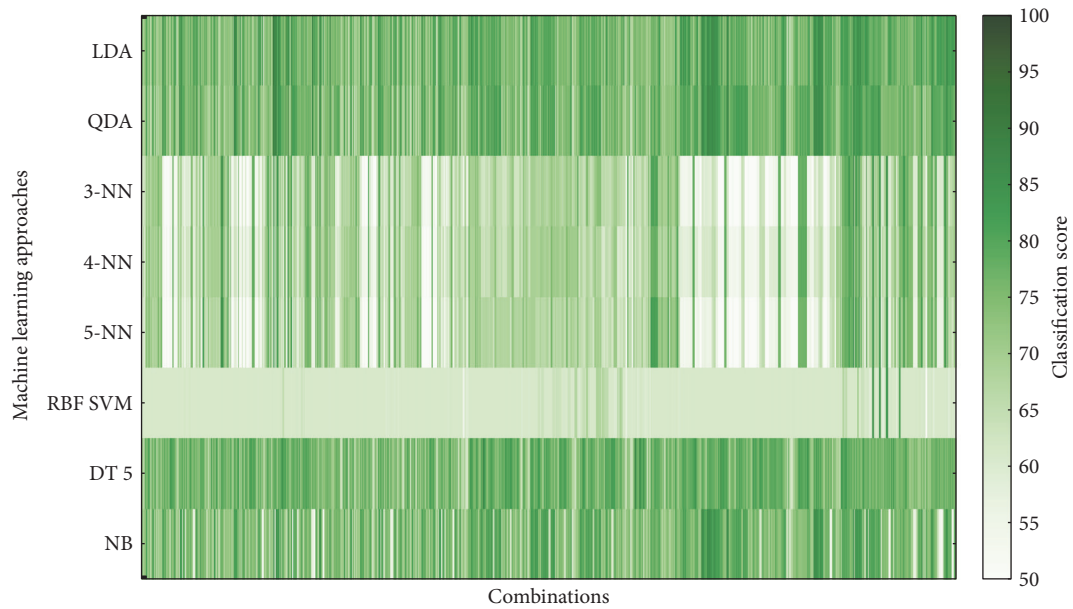


FIGURE 8: Classifier score for 4-feature combinations.

analysis. Moreover, in Figures 5–8, it can be clearly seen that approaches of discriminant analysis have more combinations with relatively high score than any other approach. Furthermore, for the support vector machine approach, only few combinations have acceptable classification score level.

It is worthy to mention that generally classification accuracy rises as the number of features in the feature set increases. For 4-feature sets, the maximum is achieved—accuracy for 5-feature sets is lower for all machine

learning approaches. It drops significantly in case of support vector machine approach.

Table 4 presents best results achieved by all machine learning approaches for 4-feature set.

Data in Table 4 shows that linear and quadratic DA not only achieve higher classification score but also have better stability of the results. Naïve Bayes classifier also has relatively high classification score and low deviation.

Among 53 studied features, 36 form combinations that have the classification score higher than 85. Table 5 presents



FIGURE 9: Classifier score for 5-feature combinations.

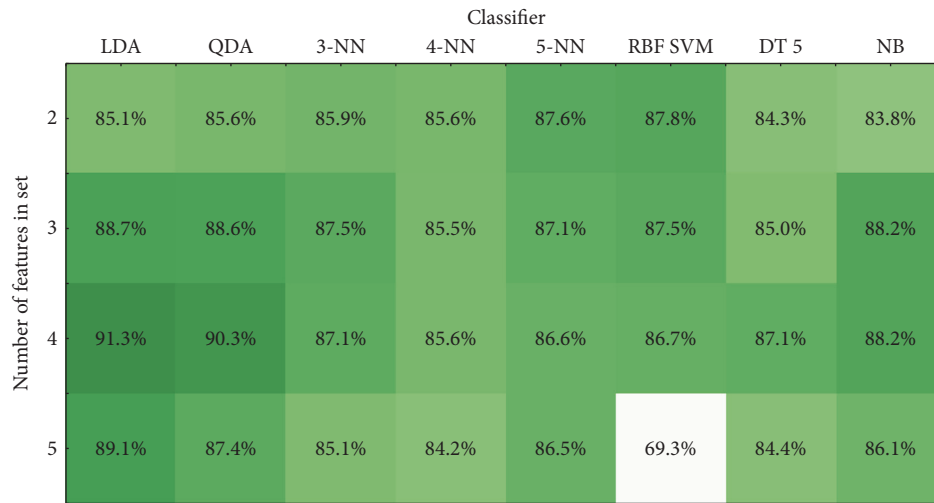


FIGURE 10: Maximal scores achieved by each learning machine approach.

occurrences of the features among the combinations. The highest occurrences are noted for different spectral features, associated with VLF spectral band, LF/HF ratio, and statistical feature heart rate.

Table 6 presents 7 features that form combinations with accuracy higher than 90%. All these combinations consist of heart rate, one feature associated with LF/HF ratio, and two features associated with VLF spectral band.

4. Discussion

For discussion purposes, a comparison of the results of the current study with results of one of the commonly used procedure, principal components analysis (PCA), was executed.

The PCA is a statistical procedure used to reveal the internal structure of the dataset [34]. In our case, features of different amplitude are used; PCA is known to be sensitive to the relative scaling of the feature dataset. Therefore, prior to the PCA application, the standardization procedure was implemented for each of 53 different features—subtraction of the mean value and after that division by the standard deviation.

Table 7 presents explained variance as well as the cumulative variance for the first 15 principal components. First 10 principal components explain 93% of the variance. Consequent principal components add 1% of the variance or less.

In order to compare results of the semioptimal search of the noncorrelated feature space with PCA, combinations of the first 10 components were consequently tested for all

TABLE 4: Best classification scores.

Score, %	Features			
Linear discriminant analysis				
91.33 ± 1.75	HR	VLF _n (Fr)	LF/HF(Fr)	VLF(wt)
90.30 ± 1.37	HR	VLF _n (Fr)	VLF(wt)	(LF/HF) _{int}
90.04 ± 1.85	HR	LF/HF(Fr)	VLF(wt)	VLF _n (wt)
90.44 ± 1.60	HR	VLF _n (Fr)	LF/HF(Fr)	SDVLF
90.11 ± 1.80	HR	LF/HF(Fr)	SDVLF	VLF _n (wt)
90.16 ± 1.61	HR	SDVLF	VLF _n (wt)	(LF/HF) _{int}
Quadratic discriminant analysis				
90.31 ± 1.71	HR	VLF _n (Fr)	LF/HF(Fr)	VLF(wt)
3-nearest neighbors				
87.14 ± 2.12	LF/HF(Fr)	SDVLF	VLF _n (wt)	W _{1/2} VLF
4-nearest neighbors				
85.56 ± 2.40	SDVLF	VLF _n (wt)	LF/HF(wt)	W _{1/2} VLF
5-nearest neighbors				
86.63 ± 1.30	HR	HF(Fr)	LF _n (Fr)	W _{1/2} VLF
Support vector machine, radial base function				
86.73 ± 2.24	IAS	RF	a ₂ LF	WVLF
Decision trees, max depth 5				
87.10 ± 3.40	IARP	LF/HF(Fr)	IAS	WLF
Decision trees, no max depth				
87.34 ± 3.08	IARP	LF/HF(Fr)	IAS	WLF
Naïve Bayes classifier				
88.17 ± 1.07	VLF(Fr)	VLF _n (Fr)	LF/HF(Fr)	W _{1/2} LF

TABLE 5: Features occurrences for classification score higher than 85%.

Features	Occurrences, %	Features	Occurrences, %
$VLF_n(\text{Fr})$	50.89	Nd	4.73
$VLF(\text{Fr})$	50.89	M_0	4.73
$VLF_n(\text{wt})$	47.93	WLF	4.73
$W_{1/2}VLF$	34.91	IARP	3.55
$LF/HF(\text{Fr})$	34.32	IC	2.96
HR	33.73	$HF(\text{Fr})$	2.96
SDVLF	30.18	$\alpha_2 LF$	2.96
$(LF/HF)_{\text{max}}$	24.26	$LF_n(\text{wt})$	2.37
$LF/HF(\text{wt})$	18.34	SI	2.37
$(LF/HF)_{\text{int}}$	18.34	$LF_n(\text{Fr})$	1.78
$W_{1/2}LF$	17.75	$\alpha_{\text{min}} LF$	1.78
H	13.61	ARI	1.78
WVLF	13.02	$HF_n(\text{Fr})$	1.78
IAS	10.65	SDHF	1.18
$VLF(\text{wt})$	7.69	IAB	0.59
RF	6.51	NN50	0.59
$\alpha_{\text{max}} LF$	5.92	$\alpha_0 LF$	0.59
M	5.33	$\alpha_{\text{max}} VLF$	0.59

TABLE 6: Feature occurrences for classification score higher than 90%.

Features	Occurrences, %
HR	100.00
$LF/HF(\text{Fr})$	62.50
$VLF(\text{wt})$	62.50
$VLF_n(\text{wt})$	50.00
$VLF_n(\text{Fr})$	50.00
SDVLF	37.50
$(LF/HF)_{\text{int}}$	37.50

TABLE 7: Dataset analysis by PCA.

Principal component	Explained variance, %	Cumulative variance, %
1	34.88	34.88
2	17.65	52.54
3	13.03	65.57
4	8.87	74.45
5	5.38	79.83
6	4.10	83.93
7	3.55	87.47
8	2.42	89.89
9	1.78	91.67
10	1.36	93.03
11	1.01	94.04
12	0.95	94.98
13	0.84	95.82
14	0.74	96.56
15	0.61	97.17

machine learning approaches using 100 repeated 5-fold cross-validation. Figure 11 presents the maximal results of classification accuracy achieved by each machine learning approach using combinations of the principal components.

Comparing the results of Figures 10 and 11, one can note that features found by the semioptimal search of the noncorrelated feature space reach higher classification accuracies than combinations of the principal components for all tested machine learning approaches.

5. Conclusions

In this work, various machine learning approaches were tested in task of the arterial hypertension diagnostics. In earlier works, the same datasets were used for investigation of the linear and quadratic DA methods [11]. The present work implies comparison of the DA methods with other machine learning approaches, like support vector machine, k -nearest neighbors, Naive Bayes, and decision trees.

The results of the current investigation showed that for the studied task, the application of the discriminant

	LDA	QDA	3-NN	4-NN	5-NN	RBF SVM	DT 5	NB
2	82.4%	81.4%	81.6%	78.2%	82.9%	67.6%	77.4%	82.2%
3	80.2%	81.5%	82.5%	77.2%	82.5%	63.6%	76.3%	79.9%
4	81.1%	83.1%	84.9%	80.7%	80.3%	60.8%	77.7%	78.8%
5	83.2%	85.5%	81.8%	78.2%	80.5%	60.9%	76.0%	78.6%
6	84.0%	84.5%	82.3%	79.6%	80.2%	60.9%	76.0%	76.8%
7	85.3%	85.0%	82.5%	78.6%	80.7%	60.9%	79.2%	76.2%
8	84.4%	82.8%	84.4%	80.9%	81.8%	60.9%	77.6%	75.7%
9	83.6%	81.1%	84.0%	80.2%	81.5%	60.8%	77.8%	73.3%
10	82.7%	81.9%	84.7%	80.0%	83.4%	61.0%	78.0%	74.0%

FIGURE 11: Scores of the PCA achieved by each learning machine approach.

analysis (linear and quadratic) revealed to be the most appropriate classifiers. These approaches have high classification score and low deviations over different realizations. A set of four features in combination seems to be the optimal number, as the classification accuracy score is higher and more consistent than those for two, three, and five features in combination.

Prevalence of the VLF and LF/HF spectral features among best combinations might indicate that sympathetic nervous system takes an important part in the initialization of the arterial hypertension and maintenance of the increased vascular tone as well as increased cardiac output. These results are in accordance with scientists' interpretation of the arterial hypertension development [35, 36].

The results of the suggested approach were compared with data set prepared by the commonly used procedure of principal component analysis. Results of the n -feature noncorrelated sets have achieved higher classification accuracies than ones based on the dataset of the selected principal components.

In future works, our research group will continue to improve results on this problem. One of the investigations that are planned is to analyze robustness of the classifiers based on multiple signals recorded simultaneously. Among the other perspective directions of future investigation is usage of the advanced neural networks [37] and genetic algorithms [38] for feature extraction and classification.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The work was supported by the Act 211 Government of the Russian Federation, Contract no. 02.A03.21.0006, and by the FCT project AHA CMUP-ERI/HCI/0046/2013.

References

- [1] T. Rosenthal and A. Alter, "Occupational stress and hypertension," *Journal of the American Society of Hypertension*, vol. 6, no. 1, pp. 2–22, 2012.
- [2] X. L. Feng, M. Pang, and J. Beard, "Health system strengthening and hypertension awareness, treatment and control: data from the China health and retirement longitudinal study," *Bulletin of the World Health Organization*, vol. 92, no. 1, pp. 29–41, 2014.
- [3] M. V. Kamath, M. Watanabe, and A. Upton, *Heart Rate Variability (HRV) Signal Analysis: Clinical Applications*, CRC Press, New York, 2012.
- [4] I. B. Ushakov, O. I. Orlov, R. M. Baevskii, E. Y. Bersenev, and A. G. Chernikova, "Conception of health: space-earth," *Human Physiology*, vol. 39, no. 2, pp. 115–118, 2013.
- [5] P. Melillo, M. Bracale, and L. Pecchia, "Nonlinear heart rate variability features for real-life stress detection. Case study:

- students under stress due to university examination,” *Biomedical Engineering Online*, vol. 10, no. 96, pp. 1–13, 2011.
- [6] F. Ebrahimi, S. K. Setarehdan, J. Ayala-Moyeda, and H. Nazeran, “Automatic sleep staging using empirical mode decomposition, discrete wavelet transform, time-domain, and nonlinear dynamics features of heart rate variability signals,” *Computer Methods and Programs in Biomedicine*, vol. 112, no. 1, pp. 47–57, 2013.
 - [7] A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, “Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 1, pp. 37–48, 2009.
 - [8] U. Parlitz, S. Berg, S. Luther, A. Schirdewan, J. Kurths, and N. Wessel, “Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics,” *Computers in Biology and Medicine*, vol. 42, no. 3, pp. 319–327, 2012.
 - [9] M. O. Mendez, J. Corthout, S. Van Huffel et al., “Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis,” *Physiological Measurement*, vol. 31, no. 3, pp. 273–289, 2010.
 - [10] V. Kublanov, A. Dolganov, and V. Borisov, “Application of the discriminant analysis for diagnostics of the arterial hypertension - analysis of short-term heart rate variability signals,” in *presented at the 4th International Congress on Neurotechnology, Electronics and Informatics*, pp. 45–52, 2016.
 - [11] V. Kublanov, A. Dolganov, and Y. Kazakov, “Diagnostics of the arterial hypertension by means of the discriminant analysis - analysis of the heart rate variability signals features combinations,” in *presented at the BIOSTEC 2017 - Special Session on Neuro-electrostimulation in Neurorehabilitation Tasks*, pp. 291–298, 2017.
 - [12] M. Malik, “Heart rate variability: standards of measurement, physiological interpretation, and clinical use,” *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.
 - [13] R. M. Baevskiy, “Analiz variabelnosti serdechnogo ritma pri ispolzovanii razlichnykh ehlektrokardiograficheskikh sistem (metodicheskie rekomendatsii) [Analysis of heart rate variability using different electrocardiographic systems (guidelines)],” *Vestnik aritmologii [Herald Arrhythmology]*, no. 24, pp. 65–87, 2001.
 - [14] F. A. Jain, I. A. Cook, A. F. Leuchter et al., “Heart rate variability and treatment outcome in major depression: a pilot study,” *International Journal of Psychophysiology*, vol. 93, no. 2, pp. 204–210, 2014.
 - [15] N. B. Haspekova, “Diagnostichestskaya informativnost monitorirovaniya variabelnosti serdechnogo ritma serdca [Diagnostic Informativeness of the heart rate variability monitoring],” *Vestnik aritmologii [Herald Arrhythmology]*, vol. 32, pp. 15–23, 2003.
 - [16] M. Adnane, Z. Jiang, and Z. Yan, “Sleep-wake stages classification and sleep efficiency estimation using single-lead electrocardiogram,” *Expert Systems with Applications*, vol. 39, no. 1, pp. 1401–1413, 2012.
 - [17] P. S. Addison, “Wavelet transforms and the ECG: a review,” *Physiological Measurement*, vol. 26, no. 5, pp. R155–R199, 2005.
 - [18] S. Mallat, *A Wavelet Tour of Signal Processing*, 2009.
 - [19] V. S. Kublanov, “A hardware-software system for diagnosis and correction of autonomic dysfunctions,” *Biomedical Engineering*, vol. 42, no. 4, pp. 206–212, 2008.
 - [20] D. Rubin, T. Fekete, and L. R. Mujica-Parodi, “Optimizing complexity measures for fMRI data: algorithm, artifact, and sensitivity,” *PloS One*, vol. 8, no. 5, 2013.
 - [21] B. B. Mandelbrot, “Multifractal power law distributions: negative and critical dimensions and other “anomalies,” explained by a simple example,” *Journal of Statistical Physics*, vol. 110, no. 3–6, pp. 739–774, 2003.
 - [22] H. E. Stanley, L. A. N. Amaral, A. L. Goldberger, S. Havlin, P. C. Ivanov, and C.-K. Peng, “Statistical physics and physiology: monofractal and multifractal approaches,” *Physica A: Statistical Mechanics and its Applications*, vol. 270, no. 1-2, pp. 309–324, 1999.
 - [23] E. A. F. Ihlen, “Introduction to multifractal detrended fluctuation analysis in Matlab,” *Frontiers in Physiology*, vol. 3, pp. 141–150, 2012.
 - [24] D. Makowiec, A. Rynkiewicz, J. Wdowczyk-Szulc, and M. Zarczynska-Buchowiecka, “On reading multifractal spectra. Multifractal age for healthy aging humans by analysis of cardiac interbeat time intervals,” *Acta Physica Polonica B Proceedings Supplement*, vol. 5, no. 1, pp. 159–170, 2012.
 - [25] D. Makowiec, A. Rynkiewicz, R. Gałaska, J. Wdowczyk-Szulc, and M. Żarczyńska-Buchowiecka, “Reading multifractal spectra: aging by multifractal analysis of heart rate,” *EPL Europhysics Letters*, vol. 94, no. 6, p. 68005, 2011.
 - [26] F. Pedregosa, G. Varoquaux, A. Gramfort et al., “Scikit-learn: machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
 - [27] G. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, vol. 544, John Wiley & Sons, 2004.
 - [28] L. E. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.
 - [29] J. A. Suykens and J. Vandewalle, “Least squares support vector machine classifiers,” *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
 - [30] P. H. Swain and H. Hauska, “The decision tree classifier: design and potential,” *IEEE Transactions on Geoscience Electronics*, vol. 15, no. 3, pp. 142–147, 1977.
 - [31] K. P. Murphy, *Naive Bayes Classifiers*, Univ. Br. Columbia, 2006.
 - [32] R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” *Ijcai*, vol. 14, pp. 1137–1145, 1995.
 - [33] P. Refaeilzadeh, L. Tang, and H. Liu, “Cross-validation,” in *Encyclopedia of Database Systems*, pp. 532–538, Springer, 2009.
 - [34] I. Jolliffe, *Principal Component Analysis*, Wiley Online Library, 2002.
 - [35] G. Parati and M. Esler, “The human sympathetic nervous system: its relevance in hypertension and heart failure,” *European Heart Journal*, vol. 33, no. 9, pp. 1058–1066, 2012.
 - [36] G. Mancina, R. Fagard, K. Narkiewicz et al., “2013 ESH/ESC guidelines for the management of arterial hypertension,” *European Heart Journal*, vol. 34, no. 28, pp. 2159–2219, 2013.
 - [37] H. B. Demuth, M. H. Beale, O. De Jess, and M. T. Hagan, *Neural network design*, Martin Hagan, 2014.
 - [38] A. Fraser and D. Burnell, *Computer Models in Genetics*, Mcgraw-Hill Book Co., New York, 1970.

